POINT PROCESSES IN EPIDEMIOLOGY

J. Gani

Stanford University

Prepared for:

Office of Naval Research

15 May 1973

Unclassified

Security Classification

DOCUMENT CONTROL DATA · R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) DEPARTMENT OF STATISTICS STANFORD, CALIF. STANFORD UNIVERS	2a REPORT SECURITY CLASSIFICATION
	2b GROUP

3. REPORT TITLE

POINT PROCESSES IN EPIDEMIOLOGY

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

TECHNICAL REPORT

5. AUTHOR(S) (Last name, First name, initial)

GANI, J.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May 15, 1973	25	25

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-67-A-0112-0053	
b. PROJECT NO. NR-042-267	#203
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	STATISTICS & PROBABILITY PROGRAM OFFICE OF NAVAL RESEARCH CODE 436 ARLINGTON, VA.

13. ABSTRACT

The paper consists of six sections: the first is devoted to chain binomial methods and their use in the statistical analysis of measles and hepatitis data. A second considers time dependent results for carrier-borne epidemics, and the use of matrix methods in computing probabilities of their final size. The third surveys the application of perturbation techniques to the general stochastic epidemic, and the estimation of infection and removal parameters in this model on the basis of smallpox data. The fourth section summarizes asymptotic results for the general stochastic epidemic when the initial populations of susceptibles and infectives are both very large. In the fifth, some recent results are outlined on the costs of epidemics; these depend on the stochastic path integral under the infective curve. Finally, a brief account is given of the analysis of space-time interactions in epidemic processes.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
epidemics: space-time interactions in cost of stochastic integral for [illegible] point processes in perturbation methods in chain binomial carrier-borne						

INSTRUCTIONS

1. **ORIGINATING ACTIVITY**: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.

2a. **REPORT SECURITY CLASSIFICATION**: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. **GROUP**: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE**: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **DESCRIPTIVE NOTES**: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S)**: Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE**: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. **TOTAL NUMBER OF PAGES**: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. **NUMBER OF REFERENCES**: Enter the total number of references cited in the report.

8a. **CONTRACT OR GRANT NUMBER**: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, &c, & 8d. **PROJECT NUMBER**: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. **ORIGINATOR'S REPORT NUMBER(S)**: Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. **OTHER REPORT NUMBER(S)**: If the report has been assigned any other report numbers (*either by the originator or by the sponsor*), also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES**: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES**: Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY**: Enter the name of the departmental project office or laboratory sponsoring (*paying for*) the research and development. Include address.

13. **ABSTRACT**: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (*TS*), (*S*), (*C*), or (*U*).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS**: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)

Unclassified

Security Classification

POINT PROCESSES IN EPIDEMIOLOGY

BY

J. GANI

TECHNICAL REPORT NO. 203
MAY 15, 1973

PREPARED UNDER CONTRACT N00014-67-A-0112-0053
(NR-042-267)
OFFICE OF NAVAL RESEARCH

Herbert Solomon, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

POINT PROCESSES IN EPIDEMIOLOGY

by

J. Gani
University of Sheffield

1. Introduction

In his basic paper on stochastic point processes Bartlett (1954) first presented some general methods for Markov processes $\{X(t)\}$, applicable to the theory of epidemics.

These methods consisted essentially in the derivation of a symbolic equation

$$\frac{\partial f}{\partial t} = \varphi\left(i\theta, \frac{\partial}{\partial i\theta}, t\right) f \qquad (1.1)$$

for the characteristic function $f(\theta) = \mathcal{E}(e^{i\theta X(t)})$ of the process $X(t)$, with $\frac{\partial}{\partial i\theta}$ operating only on f. The function $\varphi(i\theta, x, t)$ is defined as

$$\lim_{\delta t \to 0} \mathcal{E}\left\{ \frac{e^{i\theta[X(t+\delta t)-X(t)]} - 1}{\delta t} \right\} \qquad (1.2)$$

given that $X(t) = x$. An analogous equation for the probability generating function $P(z) = f(-i \ln z)$ of the process can also be derived, which corresponds to the results from the forward Kolmogorov equation. The method can be extended to vector valued Markov processes $\{\underline{X}(t)\}$

In the context of epidemics, a point process in time t will have associated with it some random vector $\underset{\sim}{X}(t)$ whose components may be one or more of the following random variables. susceptibles $R(t)$, infectives $S(t)$, carriers $Y(t)$, their locations $\{L(t)\}$ in the plane, and other related random variables such as the cost $C(t)$ of the epidemic. For convenience, epidemic processes are usually assumed to be Markovian.

It is the purpose of this review to outline some of the recent work carried out in mathematical epidemiology. In selecting the material to be discussed, several valuable developments have necessarily had to be omitted. Readers interested in a comprehensive survey up to 1967 should consult Dietz (1967); a broad sketch of current trends in epidemic theory is provided by the 17 communications in the WHO Symposium in Quantitative Epidemiology (1971). This is to appear in print very shortly.

In the present paper I have, while attempting to cover a wide range of topics, been guided in my choice largely by personal interests. My hope is that I shall succeed in making those recent developments which have attracted my attention of as much interest to my readers as myself.

The paper consists of six sections: the first is devoted to chain binomial methods and their use in the statistical analysis of measles and hepatitis data. A second considers time dependent results for carrier-borne epidemics, and the use of matrix methods in computing probabilities of their final size. The third surveys the application

of perturbation techniques to the general stochastic epidemic, and the estimation of infection and removal parameters in this model on the basis of smallpox data. The fourth section summarizes asymptotic results for the general stochastic epidemic when the initial populations of susceptibles and infectives are both very large. In the fifth, some recent results are outlined on the costs of epidemics, these depend on the stochastic path integral under the infective curve. Finally, a brief account is given of the analysis of space-time interactions in epidemic processes. We now proceed to develop these main themes.

2. Chain binomial methods

A simple advance in the discrete time treatment of small scale epidemics has been the reformulation of Greenwood and Reed-Frost chain binomial models as Markov chains. A brief outline of this was first presented by Gani (1969) and later developed in detail by Gani and Jerwood (1971). Consider a discrete time epidemic process in which the latent period of the infection is taken as the unit of time; let the random variable S_t denote the number of infected individuals just prior to time $t = 0, 1, \ldots$, who become infectious at t, while R_t is the remaining number of susceptibles. Clearly $R_t = R_{t+1} + S_{t+1}$.

If $0 < p = 1-q < 1$ is the probability of contact between any 2 individuals, then in the Greenwood model where the number of infectives at time t, when these are non-zero, is assumed not to influence the probability of infection during $(t, t+1)$,

$$\Pr\{R_{t+1} = r_{t+1} | R_t = r_t\} = \frac{r_t!}{r_{t+1}!\,(r_t - r_{t+1})!}\, p^{r_t - r_{t+1}}\, q^{r_{t+1}} \qquad (2.1)$$

$$(t = 0, 1, \ldots)$$

This clearly indicates the Markovian nature of $\{R_t\}$. For $r_0 = k$, say, the transition probability matrix of the process can be written as

$$
\begin{array}{c|cccccc}
 & r_{t+1} & & & & & \\
r_t & 0 & 1 & 2 & \cdots & k \\
\hline
0 & 1 & 0 & 0 & \cdots & 0 \\
1 & p & q & 0 & \cdots & 0 \\
2 & p^2 & 2pq & q^2 & \cdots & 0 \\
\vdots & \cdots & \cdots & \cdots & & \cdots \\
k & p^k & kp^{k-1}q & \binom{k}{2} p^{k-2}q^2 & & q^k
\end{array}
= P + Q , \qquad (2.2)
$$

where $Q = \mathrm{diag}(1, q, \ldots, q^k)$, and P is the remaining matrix with zeros in the diagonal.

The epidemic process is assumed to stop at time $T = t$ when $r_{t-1} = r_t \geq 0$. The probability of this event is given by

$$\Pr\{T = t\} = A_k' P^{t-1} QE , \qquad (2.3)$$

where $A_k' = (0, 0, \ldots, 0, 1)$ is a $(k+1)$ row vector, and E is the $(k+1)$ column vector of unit elements. From (2.3), the p.g.f. of time T to termination of the epidemic is seen to be

$$A_k'(I - \theta P)^{-1} \theta QE \qquad (0 \leq \theta \leq 1) . \qquad (2.4)$$

A similar expression can readily be obtained for the joint p.g.f. of T and the final number of infected cases at the end of the epidemic.

In the case of the Reed-Frost model, the number of infected cases s_t at time t affects infection during the interval $(t,t+1)$, and

$$\Pr\{S_{t+1} = s_{t+1},\ R_{t+1} = r_t - s_{t+1} \mid S_t = s_t,\ R_t = r_t\}$$

$$= \frac{r_t!}{s_{t+1}!\,(r_t - s_{t+1})!}\,(1 - q^{s_t})^{s_{t+1}}\, q^{s_t(r_t - s_{t+1})}. \qquad (2.5)$$

From this, it is clear that $\{S_t, R_t\}$ form a bivariate Markov chain. Techniques used in the manipulation of this chain are similar to those outlined earlier, though the matrices are now larger and more complicated.

The Greenwood model may be considered as a Markov chain imbedded in a continuous time pure death process. Though there are some similarities in their characteristics, the Reed-Frost model does not, however, correspond to the Markov chain imbedded in a simple stochastic epidemic. But the reformulation of these chain binomial models as Markov chains enables us to overcome the restrictiveness of the Greenwood and Reed-Frost infection schemes. For example, the Markov chain imbedded in the simple stochastic epidemic (for a single latent period), or in the general stochastic epidemic may serve equally well as suitable models. We may also, in the Greenwood and Reed-Frost type models, use non-homogeneous Markov chains to simulate changes in the probability p of infectious contact, possibly due to inoculation. An example of this is considered in detail in the final section of Gani and Jerwood's (1971) paper.

At a more practical level, Bailey and Alff-Steinberger (1970) have used information from basic chain binomial models to estimate parameters from an associated continuous time Markov process with

infection parameter λ. In this, the instant of infection is followed by a latent period ξ, which is normally distributed $N(\mu,\sigma^2)$, and is succeeded by an infectious period of constant length α. On the basis of Hope Simpson's measles data, assuming either a Greenwood or Reed-Frost model in households of two or three people, Bailey and Alff-Steinberger found that μ, α were of the order of 8 and 7 days respectively.

A similar analysis of Dr. K. Peterson's data on infectious hepatitis assuming a Reed-Frost model yielded values of 16 and 22 days for μ, α. It should be pointed out that in this case, differences in the cut-off point of the data resulted in sizable differences in the estimates, but a perfectly satisfactory fit is obtained from the model whether the cut-off point was 9-10 or 12-13 days.

3. Time-dependent results for carrier-borne epidemics

An interesting development in the continuous time stochastic theory of epidemics is Gillian Denton's (1971) time-dependent solution for the carrier-borne infection previously discussed by Downton (1968). Consider a closed population initially consisting of $n \geq 1$ susceptibles and $a \geq 1$ carriers at time $t = 0$. If at time $t \geq 0$ there are $0 \leq r \leq n$ susceptibles and $0 \leq s \leq n + a - r$ carriers, with the remaining $n + a - s - r$ individuals removed from the population, then the transition probabilities for the process in the interval $(t, t + \delta t)$ are given by

$\Pr\{(r,s) \to (r-1, s+1)\}$ $= \pi rs\, \delta t + o(\delta t)$	when a susceptible is infected, and becomes an undetected carrier;
$\Pr\{(r,s) \to (r-1,s)\}$ $= (1-\pi)\, rs\, \delta t + o(\delta t)$	when a susceptible is infected, detected and removed;
$\Pr\{(r,s) \to (r, s-1)\}$ $= \rho s\, \delta t + o(\delta t)$	when a carrier is detected and removed.

Here the infection rate is taken as 1, and the relative carrier removal rate as ρ, while the probability that an infective becomes an undetected carrier is $0 < \pi < 1$.

The probability generating function $P(z,w,t) = \sum_{r,s} p_{rs}(t)\, z^r w^s$, where $p_{rs}(t)$ is the probability of r susceptibles and s carriers

at time $t > 0$, given that these are respectively n and a at $t = 0$, satisfies the second order partial differential equation

$$\frac{\partial P}{\partial t} = w(w\pi + 1 - z - \pi)\frac{\partial^2 P}{\partial z\,\partial w} + (1 - w)\rho\frac{\partial P}{\partial w}, \qquad (3.1)$$

subject to the initial condition $P(z,w,0) = z^n w^a$. If we write $P(z,w,t) = \sum_{r=0}^{n} z^r f_r(w,t)$ as in Gani (1967), and take Laplace transforms

$$F_r(w,\theta) = \int_0^\infty e^{-\theta t} f_r(w,t)\,dt, \qquad \mathrm{Re}\,\theta > 0,$$

with respect to time t, (3.1) can be reduced to a set of first order partial differential equations.

These can be written in matrix form as

$$A(w)\frac{\partial F}{\partial w} + \theta F = w^a E, \qquad (3.2)$$

where $F' = F'(w,\theta) = [F_n(w,\theta), \ldots, F_0(w,\theta)]$, $E' = [1, 0, \ldots, 0]$, and $A(w)$ is given by the $(n+1) \times (n+1)$ matrix

$$A(w) = \begin{bmatrix} w(n+\rho)-\rho & & & \\ -nw(\pi w+1-\pi) & w(n-1+\rho)-\rho & & \\ & -(n-1)\,w(\pi w+1-\pi) & w(n-2+\rho)-\rho & \\ & \cdots & \cdots & \cdots \\ & & -w(\pi w+1-\pi) & w\rho-\rho \end{bmatrix}.$$

The solution to (3.2) is then obtained using Gani's (1967) method; this is

$$F(w,\theta) = \sum_{i=0}^{n+a+1} \frac{w^i}{i!} \Big(\prod_{j=0}^{i-1} B_j \Big)_{n+1} F(0,\theta) - \sum_{i=a+1}^{n+a+1} \frac{w^i}{i!} \frac{a!}{\rho} \Big(\prod_{j=a+1}^{i-1} B_j \Big)_{n+1} E. \tag{3.3}$$

Here the suffix $n+1$ indicates $(n+1) \times (n+1)$ northwest truncation of the matrix, B_j is the $2(n+1) \times 2(n+1)$ matrix

$$B_j = \begin{bmatrix} [jA^{(1)}(0) + \theta I]/\rho & j(j+1)\, A^{(2)}(0)/2\rho \\ I & 0 \end{bmatrix} \tag{3.4}$$

with $A^{(k)}(0) = \partial^k A(w)/\partial w^k \big|_{w=0}$, $k = 1, 2$, and $F(0,\theta)$ is given by

$$F(0,\theta) = \Big(\prod_{j=0}^{n+a} B_j \Big)^{-1}_{n+1} \left[\frac{a!}{\rho} \Big(\prod_{j=a+1}^{n+a} B_j \Big)_{n+1} E \right]. \tag{3.5}$$

It should be noted that all products $\prod B_j$ are taken from left to right in strict decreasing order of j, and $\prod_{k}^{k-1} B_j = I$ by definition.

If the probability of an epidemic of total size $n-r$, not counting the original a carriers, is denoted by P_{n-r} $(0 \le r \le n)$, then it follows that

$$P_{n-r} = \lim_{t \to \infty} p_{r0}(t) = \lim_{\theta \to 0} \theta F_r(0,\theta) = \lim_{\theta \to 0} \theta (F(0,\theta))_{n-r},$$

a result readily obtained from (3.5). In carrying out the necessary calculations, Gillian Denton discovered that if $P(n,a)$ denotes the matrix

$$P(n,a) = \lim_{\theta \to 0} \frac{a!\theta}{\rho} \Big(\prod_{j=0}^{n+a} B_j \Big)^{-1}_{n+1} \Big(\prod_{j=a+1}^{n+a} B_j \Big)_{n+1} ,$$

which, with the exception of the final vector E is $\lim_{\theta \to 0} \theta F(0,\theta)$, then

$$P(n,a) = \{P(n,1)\}^a . \qquad (3.6)$$

Further, the first, second, ... , (n+1)th columns of $P(n,a)$ correspond to the vectors of probabilities of the total size of the epidemic for a initial carriers and n, n-1, ... , 0 initial susceptibles respectively.

As is pointed out in the paper, these results lend considerable power to the matrix method used, particularly if one is concerned with computing values of the probabilities of the epidemic size for increasing values of the initial susceptible population up to n. These can be obtained straightforwardly from a knowledge of the matrices B_j in (3.4). The result (3.6) also holds, with the obvious minor modifications in the B_j, for the case of the general epidemic considered by Gani (1967).

4. The general stochastic epidemic

Current studies of the general stochastic epidemic have been concerned with theoretical developments involving the use of perturbation techniques, as well as more practical methods for the estimation of parameters relying on electronic computation.

An earlier application of perturbation techniques was made by Bailey (1968) to the simple stochastic epidemic involving only susceptibles and infectives. This gave asymptotically valid approximations for a large population of size N. After a change of time variable $T = Nt$, the m.g.f. $M_d(\theta,T) = e^{\theta\xi(T)}$ of the proportion $\xi(T)$ of susceptibles at time T in the deterministic case is known to satisfy the partial differential equation

$$\frac{\partial M_d}{\partial T} = \theta\left(\frac{\partial^2 M_d}{\partial\theta^2} - \frac{\partial M_d}{\partial\theta}\right), \tag{4.1}$$

the infection rate being taken as unity. The initial condition is $M_d(\theta,0) = e^{\theta/(1+\gamma)}$, where $\gamma = (N-n)/n$ is the initial ratio of infectives to susceptibles.

In the stochastic model the equivalent equation for the m.g.f. is

$$\frac{\partial M}{\partial T} = N(1-e^{-\theta/N})\left(\frac{\partial^2 M}{\partial\theta^2} - \frac{\partial M}{\partial\theta}\right).$$

Approximating to the first order in N^{-1}, this reduces to

$$\frac{\partial M}{\partial T} = \theta\left(1 - \frac{\theta}{2N}\right)\left(\frac{\partial^2 M}{\partial\theta^2} - \frac{\partial M}{\partial\theta}\right), \tag{4.2}$$

with the same initial condition $M(\theta,0) = e^{\theta/(1+\gamma)} = e^{n\theta/N}$ as before. Using only the first order perturbation on the corresponding deterministic process, Bailey was able through an eigenfunction approach to obtain useful approximations for the mean, variance and epidemic curve of the simple stochastic epidemic.

Recently Weiss (1971) has suggested an alternative perturbation method depending on the moments $\nu_r = \mathcal{E}(\xi^r)$ of the proportion ξ of susceptibles. These satisfy the differential difference equations

$$\frac{d\nu_r}{dT} = r(\nu_{r+1}-\nu_r) + N^{-1}\sum_{j=0}^{r-2}(-1)^j \binom{r}{j+2} N^{-j}(\nu_{r-j-1}-\nu_{r-j}) . \tag{4.3}$$

If ν_r is expanded in powers of N^{-1} as

$$\nu_r = \nu_r^{(0)} + \nu_r^{(1)}N^{-1} + \nu_r^{(2)}N^{-2} + \cdots , \tag{4.4}$$

and the first two terms substituted in (4.3), a simple set of differential difference equations is derived which can readily be solved. An identical technique is applicable to the general stochastic epidemic, where the population now consists of susceptibles, infectives and removals, i.e., individuals who become immune or die. In Daniels' (1971) application of perturbation techniques to the general stochastic epidemic, a slightly different approach through the cumulant generating function is used. It is shown that the first approximations to the three second order cumulants of the process satisfy the set of differential equations

$$\frac{d}{dt}\begin{bmatrix} K_{20}^{(1)} \\ 2K_{11}^{(1)} \\ K_{02}^{(1)} \end{bmatrix} = \begin{bmatrix} xy \\ -2xy \\ (x+\rho)y \end{bmatrix} + 2\begin{bmatrix} -y & -x & 0 \\ y & x-y-\rho & x \\ 0 & y & x-\rho \end{bmatrix}\begin{bmatrix} K_{20}^{(1)} \\ 2K_{11}^{(1)} \\ K_{02}^{(1)} \end{bmatrix} . \qquad (4.5)$$

Here, $x = K_{10}^{(1)}$, $y = K_{01}^{(1)}$ are the first order cumulants, which are the solutions of the deterministic equations for the susceptibles and infectives in the process, ρ being the relative removal rate. While (4.5) cannot be solved explicitly, numerical approximations are obtainable.

The validity of the general stochastic epidemic model has sometimes been called into question. Thus, an analysis of data and estimation of parameters for such a model would prove extremely valuable. In a recent paper, Bailey and Thomas (1971) have made extensive use of an IBM system 360 (Model 40) computer to analyse data from a smallpox epidemic of 30 cases in a community of 120 in SE Nigeria. After a detailed discussion of the likelihood functions based on periods between successive removals, and the number of such removals, the ML estimators of the infection and removal rates were calculated. These are respectively

$$\hat{\beta} = .00168 \pm .00047, \qquad \hat{\gamma} = .162 \pm .050 \qquad (4.6)$$

so that the relative removal rate was found to be $\hat{\sigma} = 97 \pm 22$.

To cut down the rather lengthy time taken for an exact determination of the likelihood function, a gamma type approximation was used for the distribution of periods between the (u-1)th and uth removals, when

there were s remaining infectives, given the past history of the process. Computation using this approximate method was then speeded up by a factor of 18, without seriously affecting the accuracy of the estimates. Further work is to be done to establish the value of the approximation for larger bodies of data.

5. Asymptotic results for the general epidemic

In some recent work, Nagaev and Startsev (1970) have obtained some interesting new asymptotic results for the general stochastic epidemic. Let n and m respectively be the initial susceptibles and infectives in such a model. Then, the transition probabilities of the associated imbedded Markov chain are known to be

$$\begin{aligned} \Pr\{(r,s) \to (r+1,s)\} &= \frac{n-r}{\rho + n - r} = p_r , \\ \Pr\{(r,s) \to (r,s+1)\} &= \frac{\rho}{\rho + n - r} = q_r , \end{aligned} \tag{5.1}$$

where r, s are infectives and removals respectively, ρ being the relative removal rate. The absorbing boundary for this Markov chain is the straight line $s = m+r$.

If ν_n and T_n are the final size and duration of the epidemic starting with n initial susceptibles, it is difficult to obtain explicit results for their probabilities directly from the Markov chain formulation. Clearly, if we define $\{X_j\}$ as independent random variables, geometrically distributed with probabilities $\{p_j\}$ of success, and $S_k = \sum_{j=0}^{k} X_j$, then

$$\Pr\{\nu_n > k\} = \Pr\{S_0 < m, S_1 < m+1, \ldots , S_k < m+k\}. \tag{5.2}$$

But though (5.2) is a well understood probability, its explicit expression is not easily obtained. Nagaev and Startsev have, however, succeeded

in deriving several asymptotic results for it when n and m are large; these results depend on the relation of m and n as they both tend to infinity. Typical of these are the following:

Theorem 1. The probability of the final size of a general stochastic epidemic is given asymptotically, for large n and m, such that $m = o(n)$ where a δ exists for which $0 < \delta \le \frac{\rho}{n} \le 1-\delta$ or $1+\delta \le \frac{\rho}{n} = o(m)$, by

$$\Pr(\nu_n > a_n - x b_n) = \Phi(x)\,(1 + o(1)), \tag{5.3}$$

where

$$a_n = \begin{cases} \dfrac{m}{\frac{\rho}{n} - 1} \\ n\alpha \end{cases}, \qquad b_n = \begin{cases} \left[m(\frac{\rho}{n} + \frac{\rho^2}{n^2})/(\frac{\rho}{n} - 1)^3 \right]^{1/2} & \text{if } \frac{\rho}{n} > 1 \\ \dfrac{\left[n\alpha(1 + \frac{\rho^2}{n^2})/(1-\alpha) \right]^{1/2}}{\frac{\rho}{n}(1-\alpha) - 1} & \text{if } \frac{\rho}{n} < 1. \end{cases}$$

Further, in (5.3) α denotes the solution of $\alpha + \frac{\rho}{n} \ln(1-\alpha) = 0$, and $\Phi(x)$ is the standard normal distribution function.

Theorem 2. Let $|\beta| = |m(1 - \frac{\rho}{n})| \to \infty$, where $\beta = o(m)$ and $n/m^3 = o(\beta^{-2})$, then in this case the probability of the final size of the epidemic is given asymptotically by

$$\Pr(\nu_n \ge a_n - x b_m) = \Phi(x)(1 + o(1)), \tag{5.4}$$

where $a_n = \sqrt{2nm}$, $b_n = (2n^3/m)^{1/4}$.

Theorem 3. Let $\frac{\rho}{mn} \geq \delta > 0$, then the probability of the final size of the epidemic is asymptotically of the Poisson form

$$\Pr(\nu_n = k) = \left[\frac{(nm/\rho)^k}{k!} \exp(-nm/\rho) \right] (1 + o(1)) . \qquad (5.5)$$

These results, quite apart from their analytic interest, are clearly useful in characterizing the behavior of the general stochastic epidemic in large closed populations with big initial numbers of both susceptibles and infectives.

6. The cost of epidemics

A different problem, initially broached by Becker (1970) and developed by Jerwood (1970) in a recent note and his Ph.D. thesis (1971), is the cost of epidemics. Jerwood specifies the cost of a simple stochastic epidemic in a total population of N individuals with i initial infectives, as

$$C_i = aW_i + bT_i \,, \tag{6.1}$$

where $a, b > 0$ are constants. In this, T_i is the duration of the epidemic and $W_i = \int_0^{T_i} S(t)\, dt$ the area under the stochastic path between 0 and T_i traced by the number of infectives $S(t)$ at time t. The definition (6.1), a linear combination of the duration of the epidemic and the total man-hours lost by infectives, seems realistic though possibly too simple. Using the method due to McNeil (1970), the Laplace-Stieltjes transform of the cost is found to be

$$C_i^*(\theta) = \prod_{j=1}^{N-i} \left\{1 + \frac{\theta[a(N-j) + b]}{j(N-j)}\right\}^{-1} \,, \qquad \mathrm{Re}\ \theta \geq 0, \tag{6.2}$$

where the infection rate is taken as 1 for simplicity.

For the general stochastic epidemic, also starting from i initial infectives and s susceptibles in a total population N, the cost may be written as

$$C_{is} = \int_0^{T_{is}} (aS(t) + b)\, dt = aW_{is} + bT_{is} \tag{6.3}$$

where $a, b > 0$ are again constants, $S(t)$ represents the number of infectives at time t, and T_{is}, W_{is} are respectively the duration time of the epidemic and the area $\int_0^{T_{is}} S(t)\,dt$ under the stochastic infective path between 0 and T_{is}.

If we define the joint Laplace-Stieltjes transform $\varphi_{is}(\theta_1,\theta_2)$ of (W_{is},T_{is}) as

$$\varphi_{is}(\theta_1,\theta_2) = \mathcal{E}(e^{-\theta_1 W_{is} - \theta_2 T_{is}}), \quad \text{Re } \theta_1,\ \theta_2 \geq 0, \tag{6.4}$$

then we can, again using McNeil's (1970) technique, show that these satisfy the bivariate difference equation

$$\alpha_{is}\varphi_{is} = \rho\,\varphi_{i-1,s} + s\,\varphi_{i+1,s-1}. \tag{6.5}$$

Here ρ is the relative removal rate, and $\alpha_{is}(\theta_1,\theta_2) = (\theta_1 + \theta_2 i^{-1} + s + \rho)$. We obtain formally that

$$\varphi_{is} = \prod_{j=1}^{i} \frac{\rho}{\alpha_{is}} + \frac{s}{\rho} \sum_{k=1}^{i} \varphi_{k+1,s-1} \left(\prod_{j=k}^{1} \frac{\rho}{\alpha_{js}} \right) \tag{6.6}$$

$$(i = 1, \ldots, N-s;\ s = 0, 1, \ldots, N-1).$$

Jerwood has examined in some detail the distributions of (W_{is},T_{is}), their expectations $M_{is} = \mathcal{E}(W_{is})$, $N_{is} = \mathcal{E}(T_{is})$, and hence the expectation $\mathcal{E}(C_{is})$ of the cost. He has found that (M_{is},N_{is}) satisfy the respective difference equations

$$(s+\rho)\,M_{is} = 1 + \rho M_{i-1,s} + sM_{i+1,s-1}\,,$$
$$(s+\rho)\,N_{is} = i^{-1} + \rho N_{i-1,s} + sN_{i+1,s-1}\,. \tag{6.7}$$

The first equation can be solved to give results of the form

$$M_{is} = \frac{i+s}{\rho} - \sum_{j=1}^{s} \binom{s}{j}\left(\frac{\rho}{\rho+j}\right)^{i+s-j} \beta_j\,, \tag{6.8}$$

where the β_j are themselves positive solutions of the equations

$$\sum_{j=1}^{s} \binom{s}{j}\left(\frac{\rho}{\rho+j}\right)^{s-j} \beta_j = \frac{s}{\rho} \qquad (s = 1, 2, \ldots, N)\,.$$

The $\{M_{is}\}$ are shown to have the simple bounds

$$\frac{i}{\rho} \le M_{is} \le \frac{i+s}{\rho}\,, \tag{6.9}$$

and graphical information computed for the behavior of $\{M_{is}\}$ when $i = 1, 4, 7$; $s = 3$, has demonstrated that these bounds are reasonably good.

The equation for the $\{N_{is}\}$ yields

$$N_{is} = \frac{1}{s+\rho} \sum_{k=1}^{i} \left(\frac{\rho}{s+\rho}\right)^{i-k} \left[\frac{1}{k} + sN_{k+1,s-1}\right] \tag{6.10}$$

where $N_{i0} = \sum_{k=1}^{i} k^{-1}/\rho$. Hence they can be evaluated recursively without difficulty.

A large number of interesting problems remain unsolved in this area; several of these are currently under investigation.

7. Space-time interactions in epidemics

A problem of great importance in epidemiology has been the identification of clustering among reported cases of a disease. Knox (1964 a,b) investigated methods for deciding whether leukaemia cases occurring at small distance from each other also occurred close in time; such space-time clustering would support the theory that the disease was contagious. His method was later improved upon by Barton and David (1966); a further extension by Pike and Smith (1968) allowed for limited periods of susceptibility and infectiousness.

Barton and David specified Knox's procedure in graph-theoretic terms. Let X be the number of pairs of cases from among the n reported which are found to be

(i) living within a fixed distance δ from each other

(ii) infected within a fixed time period τ of each other.

Adjacency matrices S and T of size $n \times n$ may be written such that $S_{ij} = 1$ if case j is at a distance no greater than δ from case i and is 0 otherwise, while $T_{ij} = 1$ if case j occurs within a period τ of i and is 0 otherwise. By definition, $S_{jj} = T_{jj} = 0$. The measure X of space-time clustering is then given by

$$X = \frac{1}{2} \sum_i \sum_j S_{ij} T_{ij} \, . \qquad (7.1)$$

This integer is found to be distributed approximately as a Poisson variable and can consequently be used as a test statistic. Details of other graph-theoretic structures of use in epidemiology may be found in Tautu (1970).

In Knox's method the space and time measures related to cases i and j are either 0 or 1. But S_{ij}, T_{ij} need not necessarily be restricted to these values; any reasonable measure of contact in space and time may be used. The variable X will then remain an indicator of space-time clustering; its first two moments in this general case have been obtained by Barton and David (1966). Mantel (1967) later developed a generalized regression approach to the clustering problem of which the Knox and Barton and David results are special cases. Here a statistic

$$Z = \sum_i \sum_j S_{ij} R_{ij} \qquad (7.2)$$

is defined where S_{ij}, R_{ij} are respectively functions of the space and time coordinates of cases i, j. Testing of the statistic is carried out by Monte Carlo methods, or in some cases where it appears justifiable, by the use of an approximate normal distribution for $U = (Z - \mathcal{E}(Z))/\sigma_Z$.

In a recent paper Klauber (1971) considers an extension of Mantel's method to test clustering between two sets of points. His methods are to be applied to the diagnosis of lymphatic leukaemia cases in both man and pet cats. Let (X_i, Y_i, T_i) $(i = 1, 2, \ldots, I)$ and (X_j^*, Y_j^*, T_j^*) $(j = 1, 2, \ldots, J)$ represent the space and time coordinates of two sets of observations, say on men and cats respectively. Then we may define the space and time distances between case i among men and case j among cats by

$$S_{ij} = S\{[(X_i - X_j^*)^2 + (Y_i - Y_j^*)^2]^{1/2}\}\ , \quad R_{ij} = R(|T_i - T_j^*|) \qquad (7.3)$$

where S, R are given functions, for example the reciprocals of the distances between cases, plus some adjustable constant.

The statistic (7.2) may still be used; it can be compared to its randomization distribution by assuming one set of points fixed, and the coordinates of the other randomly permuted, or both sets may be taken as random. The mean and variance of Z can be found for both randomization models, and in some instances the approximate standard normal variable $U = (Z - \mathcal{E}(Z))/\sigma_Z$ may again be used as the test statistics. As might be expected, this approximation improves for larger samples. Two examples are given of tests based on empirical data; one of them is concerned with lymphoma cases for 117 cats and 93 dogs in California, where no statistical significance at a reasonable level was found for clustering.

It will be apparent from this brief sketch of current work in epidemiology that the field is in a state of rapid development. The intrinsic interest of the problems considered, and the liveliness of research in this area will, I hope, encourage more probabilists and statisticians to contribute to it in future.

References

Bailey, N.T.J. (1968) A perturbation approximation to the simple stochastic epidemic in a large population. Biometrika 55, 199-209.

Bailey, N.T.J. and Alff-Steinberger, Cynthia (1970) Improvements in the estimation of the latent and infectious periods of a contagious disease. Biometrika 57, 141-153.

Bailey, N.T.J. and Thomas, A.S. (1971) The estimation of parameters from population data on the general stochastic epidemic. Theoret. Pop. Biol. 2,

Bartlett, M.S. (1954) Processus stochastiques ponctuels. Ann. Inst. H. Poincaré 14, 35-60.

Barton, D.E. and David, F.N. (1966) The random intersection of two graphs. Research Papers in Statistics (F.N. David, Ed.) John Wiley, New York.

Becker, N. (1970) Mathematical models in epidemiology and related fields. Ph.D. thesis, Sheffield University.

Daniels, H.E. (1971) A note on perturbation techniques for epidemics. WHO symposium on Quantitative Epidemiology, Moscow 23-27, November 1970. Adv. Appl. Prob.

Denton, Gillian M. (1971) On the time-dependent solution of Downton's carrier-borne epidemic. Manchester-Sheffield School of Probability and Statistics Research Report, July 1971.

Dietz, K. (1967) Epidemics and rumours--A Survey. J.R. Statist. Soc. A 130, 505-528.

Downton, F. (1968) The ultimate size of carrier-borne epidemics. Biometrika 55, 277-289.

Gani, J. (1967) On the general stochastic epidemic. Proc. Fifth Berkeley Symposium, Vol. 4, 271-279.

Gani, J. (1969) A chain binomial study of inoculation in epidemics. Bull. I.S.I. 43 (2), 203-4.

Gani, J. and Jerwood, D. (1971) Markov chain methods in chain binomial epidemic models. Biometrics 27,

Jerwood, D. (1970) A note on the cost of the simple epidemic. J. Appl. Prob. 7, 440-443.

Jerwood, D. (1971) Cost of epidemics. Ph.D. thesis, Sheffield University.

Klauber, M.R. (1971) Two-sample randomization tests for space-time clustering. Biometrics 27, 129-142.

Knox, G. (1964a) Epidemiology of childhood leukaemia in Northumberland and Durham. Brit. J. Prev. Med. 18, 17-24.

Knox, G. (1964b) The detection of space-time interactions. Appl. Statist. 13, 25-9.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. Cancer Res. 27, 209-20.

McNeil, D.R. (1970) Integral functionals of birth and death processes and related limiting distributions. Am. Math. Statist. 41, 480-485.

Nagaev, A.N. and Startsev, A.N. (1970) Asymptotic analysis of a stochastic epidemic model. Teor. Veroyat. Primen. 15, 97-105.

Pike, M.C. and Smith, P.G. (1968) Disease clustering: a generalization of Knox's approach to the detection of space-time interactions. Biometrics 24, 541-536.

Tautu, P. (1970) Structural models in epidemiology: an introductory investigation. Paper presented to the WHO Symposium on Quantitative Epidemiology, Moscow 23-27 November 1970.

Weiss, G.H. (1971) On a perturbation method for the theory of epidemics. WHO Sympsoium on Quantitative Epidemiology, Moscow 23-27, November 1970. Adv. Appl. Probl.

WHO Symposium on Quantitative Epidemiology. Moscow 23-27 November 1970. (1971) Adv. Appl. Prob.